

AD-A043 047

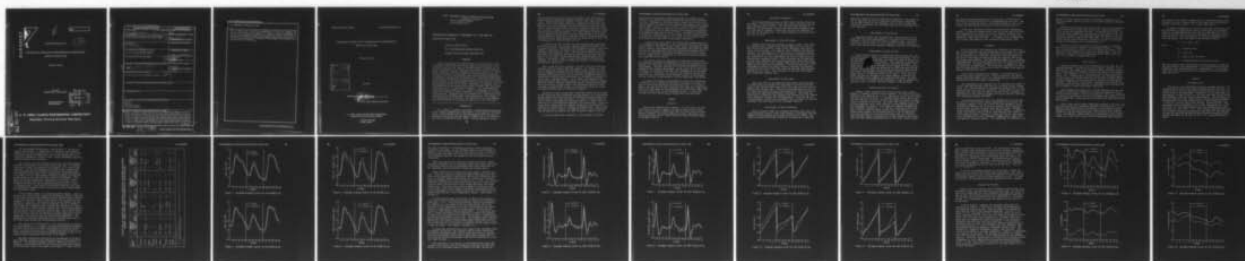
HUMAN ENGINEERING LAB ABERDEEN PROVING GROUND MD
PHYSIOLOGICAL CORRELATES OF PERFORMANCE IN A LONG DURATION REPE--ETC(U)
JUN 77 N J CARRIERO
HEL-TM-21-77

F/G 5/10

UNCLASSIFIED

NL

1 OF 1
AD
A043 047



END
DATE
FILMED
9 -77
DDC

AD A043047



Q

AD

Technical Memorandum 21-77

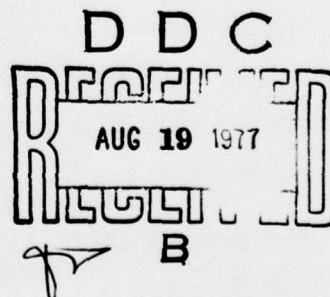
12

PHYSIOLOGICAL CORRELATES OF PERFORMANCE IN A LONG DURATION
REPETITIVE VISUAL TASK

Nicholas J. Carriero

June 1977
AMCMS Code 611102.74A0011

Approved for public release;
distribution unlimited.



U. S. ARMY HUMAN ENGINEERING LABORATORY

Aberdeen Proving Ground, Maryland

AD NO.
DDC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Memorandum 21-77	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER 9
4. TITLE (and Subtitle) PHYSIOLOGICAL CORRELATES OF PERFORMANCE IN A LONG DURATION REPETITIVE VISUAL TASK.	5. TYPE OF REPORT & PERIOD COVERED Final rept.	
7. AUTHOR(s) Nicholas J. Carriero John	8. CONTRACT OR GRANT NUMBER(s)	
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Human Engineering Laboratory Aberdeen Proving Ground, MD 21005	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS AMCMS Code 61TT02.74A0011	
11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE June 1977	13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) HEL-TM-21-77	15. SECURITY CLASS. (of this report) Unclassified	
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE		
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited. 12 28p.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Vigilance Visual Task Physiological Correlates Long-Term Performance		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This study examined the effectiveness of heart activity (HR), respiration (RESP), muscle activity (EMG), skin conductance (SCL), and brain wave activity (EEG) as discriminators of correct vs. incorrect performance in a repetitive visual task of approximately 2 hours' duration. Separate analyses were made of the data to distinguish the operation of task difficulty from performance accuracy. In addition, both of the analyses were repeated using standard score transforms of the raw data to compensate for individual differences. An interactive statistical design was employed to assess the differential changes of the physiological variables with accuracy over time. This (Continued)		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

172 850

4B

20. ABSTRACT (Continued from front)

→ design proved to be of crucial importance in assessing this relationship since the accuracy main effect for RESP, EMG, and HR was nonsignificant in all four data treatments. The accuracy-by-time interactions were significant in a number of instances and established the efficacy of these parameters as discriminators of performance adequacy. Additionally, the standard score transforms proved essential to establishing these relationships when the variance in task difficulty was eliminated. The implication of these findings for the development of an alertness indicator is also discussed. ↗

PHYSIOLOGICAL CORRELATES OF PERFORMANCE IN A LONG DURATION
REPETITIVE VISUAL TASK

Nicholas J. Carriero

ACCESSION for	
NTIS	White Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	Avail. and/or SPECIAL
A	

June 1977

APPROVED:

John D. Weisz
JOHN D. WEISZ
Director

U.S. Army Human Engineering Laboratory

U. S. ARMY HUMAN ENGINEERING LABORATORY
Aberdeen Proving Ground, Maryland 21005

Approved for public release;
distribution unlimited.

From: VIGILANCE: Theory, Operational Performance and
Physiological Correlates
Edited by Robert R. Mackie
[Plenum Press, 1977]

PHYSIOLOGICAL CORRELATES OF PERFORMANCE IN A LONG DURATION
REPETITIVE VISUAL TASK

Nicholas John Carriero

U.S. Army Human Engineering Laboratory

Aberdeen Proving Ground, Maryland, USA

ABSTRACT

This study examined the effectiveness of heart activity (HR), respiration (RESP), muscle activity (EMG), skin conductance (SCL), and brain wave activity (EEG) as discriminators of correct vs. incorrect performance in a repetitive visual task of approximately 2 hours' duration. Separate analyses were made of the data to distinguish the operation of task difficulty from performance accuracy. In addition, both of the analyses were repeated using standard score transforms of the raw data to compensate for individual differences. An interactive statistical design was employed to assess the differential changes of the physiological variables with accuracy over time. This design proved to be of crucial importance in assessing this relationship since the accuracy main effect for RESP, EMG, and HR was nonsignificant in all four data treatments. The accuracy-by-time interactions were significant in a number of instances and established the efficacy of these parameters as discriminators of performance adequacy. Additionally, the standard score transforms proved essential to establishing these relationships when the variance in task difficulty was eliminated. The implication of these findings for the development of an alertness indicator is also discussed.

INTRODUCTION

In an effort to improve performance in monitoring and vigilance tasks, various attempts have been made to assess operator efficiency level. One technique that has been employed involves the use of a secondary task and giving the operator feedback as to the effectiveness of his performance (Buck, 1968). Another approach involves

monitoring one or more physiological parameters and activating an alerting device when these parameters deviate from some criterion state. Kennedy and Travis (Kennedy & Travis, 1947, 1948; Travis & Kennedy, 1947, 1949) demonstrated in a laboratory environment that EMG activity that originated from the forehead just above the supra-orbital ridges could be used to activate a signal (light and/or buzzer) to alert a subject (*S*) when muscle tension had fallen to a level associated with poor performance. In a later report, Kennedy (1953) lists a number of difficulties that were encountered in an attempt to implement this system in an airplane cockpit. Among them are the problems of 60-Hz interference and the low signal-to-noise ratio that is encountered in amplifying EMG activity.

In addition to EMG, a number of other physiological parameters--brain wave activity, circulatory changes, pupillary response, electro-dermal activity, respiration, and biochemical activity--have been studied as indicators of attention and alertness (Kahneman, 1973; Kornblum, 1973; O'Hanlon, 1970). The results of these studies have been both frustrating and intriguing--frustrating in that there appears to be no simple relationship between these parameters and attention; intriguing in that, in spite of sometimes contradictory findings, the results offer a promise that there is a relationship waiting to be teased out.

Part of the difficulty appears to be due to the fact that physiological activity varies with task demands. For example, a large body of empirical evidence indicates that cardiac *deceleration* generally accompanies attention to the external environment, while on the other hand cardiac *acceleration* typically accompanies motivated inattention to the external environment or attention to internal activity (e.g., problem solving). As a result, experimental tasks which exclusively make either internal or external attention demands would be likely to produce conflicting results, whereas those tasks that involve both to approximately the same degree would be likely to produce nonsignificant results if analyses were based on values averaged over both types of attentional demands.

Research in this area has been stimulated largely by the studies of the Laceys and their colleagues and consequently this relationship between direction of attention and differential cardiac response has become known as the Lacey Hypothesis. It should be noted, however, that in spite of the empirical agreement there is a theoretical controversy as to the mechanism underlying this relationship. Lacey and Lacey (1974) propose that the changes in HR play a facilitative/inhibitory role with regard to stimulus intake/rejection. Obrist, Howard, Lawler, Galosy, Meyers, and Gaebelin (1974), on the other hand, view HR changes as primarily an index of general somatic activity.

A second element that contributes to the problem of finding

unequivocal relationships between physiological variables and attention is the marked individual differences found in physiological activity. Although, as indicated above, cardiac *deceleration* usually accompanies attention to the external environment, Lacey (1959) reports that 2 out of 15 *Ss* in a study of visual attention responded with cardiac *acceleration* and that 4 out of 16 *Ss* (3 at a statistically significant level) showed cardiac acceleration in an auditory discrimination task (Lacey & Lacey, 1974). Individual differences also manifest themselves in the wide range of values typically found in measuring a given variable, e.g., basal skin resistance ranges from about 10K to 500K ohms in the general population.

A third factor that affects the relationship between physiological parameters and attention is the general physiological condition of the subject. Williams, Granda, Jones, Lubin, and Armington (1962) found a positive relationship between performance decrement and theta activity in sleep-deprived *Ss* who had bimodal EEG frequency distributions (an individual difference variable) but only one of these five *Ss* showed such a relationship in the baseline period preceding the sleep deprivation period.

The present study was undertaken to determine to what extent physiological parameters differentially co-vary with accuracy of performance in a task that involved both external and internal attention demands. The task required *Ss* to scan a 4x4 letter array (attending to the external environment) and to keep track of the number of letters that appeared more than once (attending to the internal environment). In essence, the focus of the study was on the interaction (in the statistical sense) between accuracy of performance and the variation in physiological activity during the task period.

To reduce the variance attributable to individual differences, a standard score transform was applied to the data. An effort was made to control the physiological condition of the *Ss* by running them at the same time of day, requesting that they eat a light lunch before the experiment and asking smokers to refrain from smoking for at least 2 hours prior to the experiment period. *Ss* were also screened for gross physiological abnormalities and drug usage.

METHOD

Subjects

Twelve paid volunteer male subjects, 17 to 25 years old, were recruited from a local community college to serve in this study. All *Ss* were screened to insure that they were free from gross physiological defects, not taking medicine or drugs, and had at least 20/20 correctable vision. Data from two *Ss* were discarded because of equipment failure.

Recording Instruments

The following physiological variables were recorded during the experiment: skin resistance, heart rate, muscle potential, respiration, and brain wave activity. All variables were recorded on a Honeywell 1612 Visicorder and digitized and stored on magnetic tape via a DEC Lab 8/E computer. Brain wave activity was digitized at the rate of 100 samples/sec. All other parameters were digitized at the rate of 10/sec.

Measurement of Skin Resistance

Beckman Skin Electrodes (650418), Ag-Ag-Cl, with a 10-mm diameter adhesive mask, placed on the middle segment of the volar surface of the first and third finger of the left hand, were used to record skin resistance. K-Y surgical jelly (Johnson's Professional Products, Co., New Brunswick, NJ) was used as an electrode medium. The skin surface at these sites (and all other electrode sites with the exception of EEG scalp placements) was first cleaned by brisk rubbing with isopropyl alcohol. Skin resistance was recorded with an Electronic Laboratory Skin Resistance Analyzer Model 308-B which measures both galvanic skin response (GSR) and basal skin resistance (BSR) from the same set of electrodes. The device furnished a 7.85 microamperes constant excitation current, providing an equivalent current density of $10 \mu\text{a}/\text{cm}^2$ when used with the mask and electrodes employed in this study. Skin resistance values were transformed into conductance measures (SCL).

Measurement of Heart Rate

Heart rate (HR) was recorded from three Grass gold electrodes--two were placed on the upper rib cage about 10 cm to the left and right of the sternum (the ground on the left side) and the third was affixed over the apex beat. Grass EC-2 electrode paste was used as the conducting medium for these electrodes and for those employed in measuring the remaining physiological parameters. The signal was processed by a Honeywell high frequency EKG channel which was coupled with a Honeywell Accudata 131 Cardiotachometer. Both the EKG and cardiotachometer outputs were recorded on the Visicorder but only the tachometer data were digitized for subsequent analysis.

Measurement of Muscle Potential

Muscle potential (EMG) was recorded from three Grass gold electrodes which were placed on the right forearm in accordance with the standard forearm extensor pattern described by Davis (1959). The signal was processed through a Honeywell Accudata 108 AC amplifier

which was coupled with an Accudata 109 DC amplifier. The output of the DC amplifier was fed to a Honeywell Physiological Integrator. The raw EMG signal and the integrated value of the rectified EMG signal were recorded on the Visicorder but only the integrator data were digitized.

Measurement of Respiration

Respiration (RESP) was recorded from a mercury-filled strain gauge that was placed approximately 2.5 cm below the rib cage and fastened to the *S*'s back with strips of adhesive tape. The strain gauge was stretched from 7 to 10 cm beyond its relaxed extension. This signal was processed through a Honeywell Accudata 137A amplifier.

Measurement of Brain Activity

Brain activity (EEG) was recorded from the occipital area, employing the bipolar method, via three Grass gold electrodes. Two of the electrodes were placed about 5 cm above the inion at points 3 cm to the right and left of the midline. The scalp area at these locations was cleaned with acetone prior to electrode placement. The third electrode served as a ground and was placed over the left mastoid process. The signal was processed by a MED Associates EEG amplifier, passed through a MED Associates 60-Hz active notch filter and finally sent through a Krohn-hite Model 3750B filter with low cutoff set at 1 Hz and high cutoff set at 50 Hz. An on-line frequency analysis of consecutive 10-second segments of EEG activity was performed on the computer and outputted to DEC tape. The analysis consisted of determining the amplitudes of the 50 integral frequencies, from 1 to 50 Hz, that were present in the waveform.

Stimulus Material and Control

Stimulus slides consisted of 4x4 letter arrays with varying numbers of randomly selected letters repeated. Five levels of letter repetition were employed (0, 2, 3, 4, and 5) and a total of 200 slides were used in the test series. A repeated letter was one that occurred more than once, but was counted as one repeat no matter how many times it occurred in the array. The series was composed of 40 slides of each repetition level--each slide was a unique random selection and arrangement of letters. The slides were randomly arranged into four different sequences (A, B, C, and D), consisting of 50 slides each, subject to the following restrictions: Each sequence was to contain equal numbers of each repetition level and no repetition level was permitted to occur more than three times in succession. Four different orders of presentation of these sequences were developed with each sequence appearing once in the first, second, third,

and fourth presentation position to insure that any differential difficulty of the blocks would be counterbalanced across time. The first *S* was assigned to order 1, the second to order 2, etc.; the process was then repeated starting with the fifth and ninth *Ss*.

Slides were controlled by Massey Dickinson (M-D) programming modules. Slides were advanced once every 10 seconds and slide trays were changed after 75 slides (25 sets of stimulus, response, and confidence slides [see Procedure]). Two trays or 50 stimulus presentations constituted one block. *Ss*' responses were recorded manually as they were heard over the intercom in the control room.

Procedure

The *Ss* reported to the laboratory at 1300 hours and after the EKG and BSR electrodes and the respiration strain gauge were attached they were seated in a comfortable upholstered chair where the rest of the electrodes were attached. During this time the purpose of the various electrodes was explained and any questions concerning the process were answered. The experimental chamber was a shielded, sound-dampened, and air-conditioned room which was isolated from the equipment room. Overhead illumination in the *S*'s chamber was kept on throughout the experiment. The slides were projected on a screen that was approximately 2 m from the *S* with its center point approximately 60 cm above eye level.

While system calibration was underway, a prerecorded tape of instructions was played for the *S*. After calibration, the experimenter determined whether the *S* was experiencing any discomfort and adjusted any items that might be bothersome.

The experimental task required the *S* to determine the number of letters that appeared more than once in a 4x4 array on the stimulus slide. The *S* was instructed to withhold his response until a slide containing the letter "R" was projected. Finally when a blank slide appeared, the *S* was asked to give an estimate of his confidence in the accuracy of his previous judgment on a scale from 0 to 100% with 100% indicating maximum confidence.

To increase motivation, the *S* was informed that he could earn up to 2½¢ for each correct response depending on his accuracy and confidence judgment--if he was correct and 100% confident, he was told he would receive 2½¢; if he was incorrect and 50% confident, he would lose 1¼¢ from his bonus money and that other confidence values would affect his earnings in a similar fashion. In addition, he was told that no money would be taken from him if he ended up with a negative bonus. Actually, *Ss* were paid 2½¢ for each correct score irrespective of their confidence judgment and no deductions were made for incorrect answers. This subterfuge was employed in an

effort to obtain a greater spread in confidence judgment since in previous studies some *Ss* gave only two or three different confidence judgments.

Before the test series was started, the *S* was given 20 practice trials during which he was given the correct response immediately after he gave his confidence estimate. During the test series, the *S* was not given feedback nor was he aware of the length of the series or the duration of the experiment other than the fact that the experiment would be completed by 1800 hours. No clock was present in the subject room nor was the *S* permitted to wear his wristwatch during the performance of the experimental task.

Each trial in the test series lasted 30 seconds (10 seconds each for the stimulus, response, and confidence phases). A single tray, consisting of 25 trials, took 12.5 minutes to present. The duration of the complete test series was approximately 2 hours (1 hour and 40 minutes for the presentation of the eight slide trays and 2 to 3 minutes to change each tray).

Data Analysis

Averaged response curves were determined for the RESP, HR, SCL, and EMG variables. The curves are of 30 seconds' duration, starting with the .5th second of the confidence phase, extending through the stimulus phase, and ending with the 10th second of the response phase. This sequence is *not* the sequence used in the experimental trials (i.e., stimulus phase, response phase, and confidence phase) but was chosen so that the effect of pre- and post-stimulus changes in these parameters could be evaluated in relation to performance accuracy. This choice, of necessity, eliminated 8 of the 200 trials from this analysis since there was no confidence phase preceding the first stimulus in each of the eight slide trays.

Each parameter was sampled at .5-second intervals and the mean values for correct and incorrect trials were determined for each *S* at each of these points. Data for a particular trial were classified as correct or incorrect on the basis of whether the stimulus was correctly identified. Group analyses are based on the mean data derived from each *S*.

Since slides with 0-letter repetitions produced an 8% error rate while those with five repetitions produced a 70% error rate, it is possible that analyses of the differences in physiological values between correct and incorrect trials, based on the total stimulus population, may have been affected by this difference in difficulty level. Accordingly, another analysis was made based solely on performance on slides with 3- and 4-letter repetitions which produced

error rates of 44% and 55% respectively (76 such presentations were available for each S because four of the eight trays started with a stimulus slide with either 3- or 4-letter repetitions).

In order to compensate for the wide individual differences that are encountered in physiological data, further analyses were made based on a transformation that put each S on the same scale both in terms of the mean value for a given parameter and deviations from this mean value. The transformation is called a standard score by Dixon and Massey (1957) and employs the following formula:

$$Z_i = 50 + 10(X_i - \bar{X})/s$$

where

Z_i = standard score

X_i = raw score

\bar{X} = mean of the raw scores

s = standard deviation of the raw scores

Both the analyses based on the total stimulus population (STIM[ALL]) and those based on the subpopulation of 3- and 4-letter repetitions (STIM[3,4]) were repeated using this transformation. In the subsequent discussion raw score data are referred to as RS and standard score data as SS.

RESULTS

Performance Data

Since two S s were eliminated because of equipment failure, full counterbalancing of the four random sequences was not achieved. Sequences B and C, however, were equally represented in the first and second halves of the experiment. Sequence A, on the other hand, was represented six times in the first half and four times in the second, while Sequence D had the reverse representation. The differences in the mean number of errors between first- and second-half performance for Sequences A and D were nonsignificant. Mean differences were .50 and .92, and $t(8)$ values were .091 and .236 respectively. The performance data for Sequences A and D were combined and the resulting first- vs. second-half analysis was also nonsignificant; mean difference was .40, $t(18) = .127$. These results indicated that there was no reason to reject the hypothesis that effective counterbalancing had been achieved and the data were therefore examined for performance decrement on the basis of first- vs. second-half performance.

For the STIM(ALL) analysis there was a nonsignificant decrease in error production between first-half ($\bar{X} = 42$) and second-half ($\bar{X} = 41$) performance, $t(9) = .293$, $p > .05$. The STIM(3,4) analysis showed an increase in errors from the first half to the second half ($\bar{X} = 18.2$ and 21.2 respectively), but this difference was not significant, $t(9) = 1.456$, $p > .05$.

Confidence Data

There was a slight, but nonsignificant, increase in the Ss' confidence from the first to the second half of the experiment ($\bar{X} = 83.7$ and 86.4 respectively), $t(9) = 1.286$, $p > .05$. Similar analyses were made for correct and incorrect trials with the following results: During correct trials, first-half mean confidence was 85.3 vs. 88.0 for the second half, $t(9) = 1.144$, $p > .05$; for incorrect trials, means were 81.4 and 84.7 respectively, $t(9) = 1.381$, $p > .05$. The data indicate that time did not significantly influence the Ss' confidence estimates.

The data do indicate that the Ss , as a group, were significantly more confident when they were correct than when they were incorrect, $\bar{X} = 86.7$ and 83.1 respectively, $t(9) = 3.513$, $p < .05$. On an individual basis, however, only 4 of the 10 Ss were significantly more confident when they were correct and 1 S was more confident when he was incorrect, although at a nonsignificant level.

Physiological Data

The data were screened for artifacts prior to the classification of trials as correct or incorrect and before any analyses were performed to eliminate any experimenter biasing of results.

Since inadequate respiration data was obtained from the seventh S during the last half of the experiment, his contribution to the analysis of this parameter is based on first-half data only.

Analyses Based on Averaged Response Curves

Separate analyses of variance (ANOVA) were computed for each phase of the experimental trial (confidence, stimulus, and response phase) for each of the data treatments (STIM[ALL]-RS, STIM[ALL]-SS, STIM[3,4]-RS, and STIM[3,4]-SS). Each of the parameters (HR, RESP, SCL, and EMG) for which averaged response curves were developed was analyzed in this manner. A total of 48 ANOVAs were computed for this part of the analysis using a two-factor within-subjects ANOVA. The first factor was the accuracy dimension with two levels--correct vs. incorrect--while the second factor was the time dimension with

20 levels--20 points measured at .5-second intervals throughout each 10-second phase.

The accuracy-by-time interaction (AxT) was the item of major interest in these analyses because the task demanded both attention to the external environment (scanning the letter arrays) and to the interval environment (keeping a tally of the number of letters that appeared more than once). This dual demand coupled with the Lacey Hypothesis suggested that averaged HR response curves should reflect differences in accelerative and decelerative peaks as a function of accuracy and the task demands that varied with time, particularly within the stimulus phase. The other parameters were treated in the same fashion to determine to what extent they paralleled HR activity. A group of *a priori* comparisons was planned for each point in time within a phase to pinpoint the location of significant interactions. These comparisons entailed the computation of a critical difference for each of the interactions. The method detailed by Snedecor and Cochran (1967) was employed. In computing these values the two-tailed *t* value for 120 *df* at the .05 level was used, although in all cases more than 120 *df* were available for the significance test--this yielded a slightly conservative test of the differences.

The results of these ANOVAs will be discussed in conjunction with Figures 1 through 16. In these figures, the vertical lines divide the curves into three sections corresponding to the confidence, stimulus, and response phases respectively. The HR curves were derived from cardiometer data. The operational characteristics of this instrument are such that its output lags the actual beat-to-beat heart rate changes by approximately 1 second. As a consequence, what the computer plots show as occurring at, say, the 10th second, for example, actually occurred at the 9th second.

Before examining the results in terms of the AxT interactions, there are several general features of the averaged response curves that merit comment. The HR and RESP curves show a correspondence that would be anticipated from the intimate interrelationship of these variables that has been demonstrated in previous research. In particular, HR peaks lag respiration peaks even when the cardiometer lag is compensated for. Both of these parameters show greater peak activity in the confidence and response phases as would be expected from the verbal activity (*Ss'* reports) required during these periods. The SCL curves also show the typical increments that generally follow respiratory peaks. Integrated EMG activity is maximum in the anticipatory period preceding the *S's* exposure to the stimulus slide. The congruence of these results with the findings of previous research indicates that satisfactory measurement of these physiological parameters was achieved in the experiment.

In the subsequent discussion, F-ratios with $p < .05$ are considered significant. It should also be noted that even though an interaction may not be significant, it is possible for some of the differences between the means that make up the interaction to be significant (Snedecor & Cochran, 1967). These differences have been reported wherever they occurred.

The results of the AxT interaction and the critical difference analyses are presented in the table on the next page. An examination of this table in conjunction with Figures 1 to 4 reveals the following: All the AxT interactions for HR in the stimulus phase with the exception of the STIM(3,4)-RS analyses were significant. For STIM(ALL)-RS the data display significantly greater HR deceleration early in the stimulus phase followed by greater acceleration at the middle of this period during incorrect performance. The STIM(3,4)-SS data display the same pattern (see Figures 1 and 4). While the greater significant acceleration is also seen in the STIM(ALL)-SS analyses, the increased deceleration that is present is nonsignificant. The STIM(3,4)-RS analysis shows only one point of significant difference which occurs at the 10.5th second during the deceleratory phase.

In the stimulus phase, the following pattern emerges: Analyses based on raw scores reveal significant differences almost exclusively in the condition where the combined stimulus population was examined--a condition in which stimulus difficulty as well as accuracy of performance is operating. The analyses based on the STIM(3,4)-RS subpopulation, where accuracy alone is operating, failed to produce a significant AxT interaction and indicated only one point of significant difference. These findings suggest that the raw score approach is insensitive to HR covariation with accuracy. The standard score method, however, while revealing much the same pattern as was found in the raw score analysis for the total stimulus population, also demonstrates this pattern with the STIM(3,4)-SS population. The standard score technique thus possesses the desired sensitivity to the relationship between HR and performance accuracy.

The pattern of the differences is also illuminating in terms of the Lacey Hypothesis. It suggests that during incorrect performance, Ss overly attend to the external environment (scanning the letter array) in the initial part of the stimulus phase and follow this by overly attending to the internal environment (keeping a tally of repeated letters) during the middle part of this phase.

One other significant interaction was encountered in the HR results and this occurred in the confidence phase of the STIM(3,4)-SS analysis. The critical-difference analysis indicates that this interaction is associated with the two accelerative peaks that appear in the incorrect curve in this phase (see Figure 3). There is only

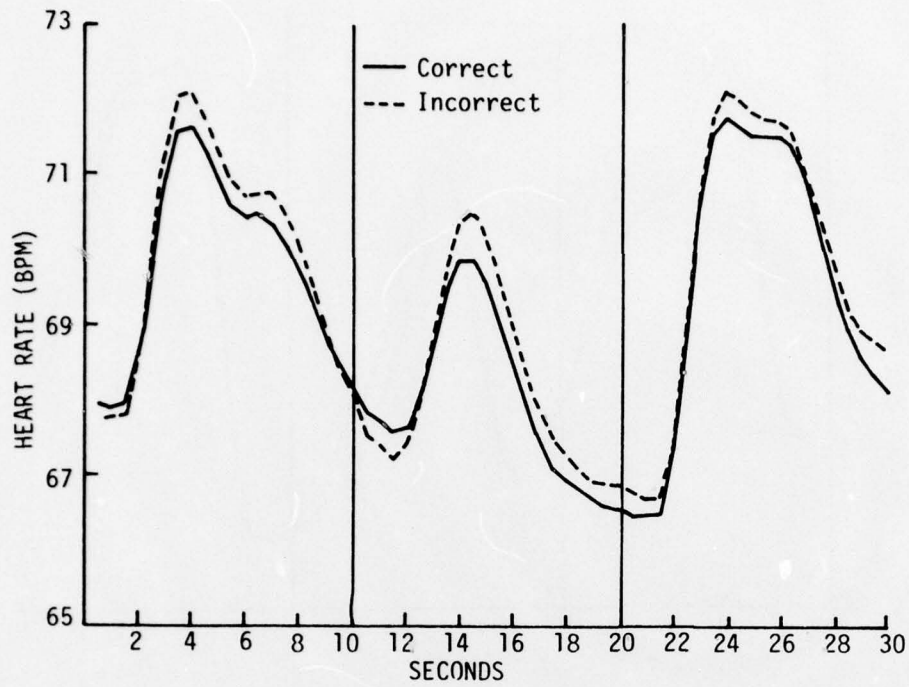


Figure 1. Averaged response curves for HR STIM(ALL)-RS.

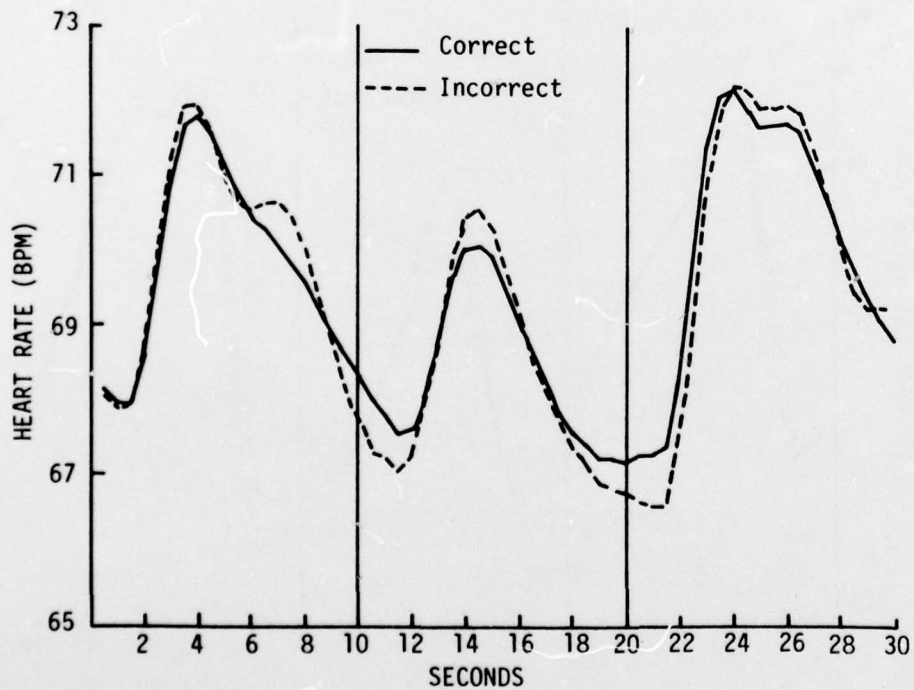


Figure 2. Averaged response curves for HR STIM(3,4)-RS.

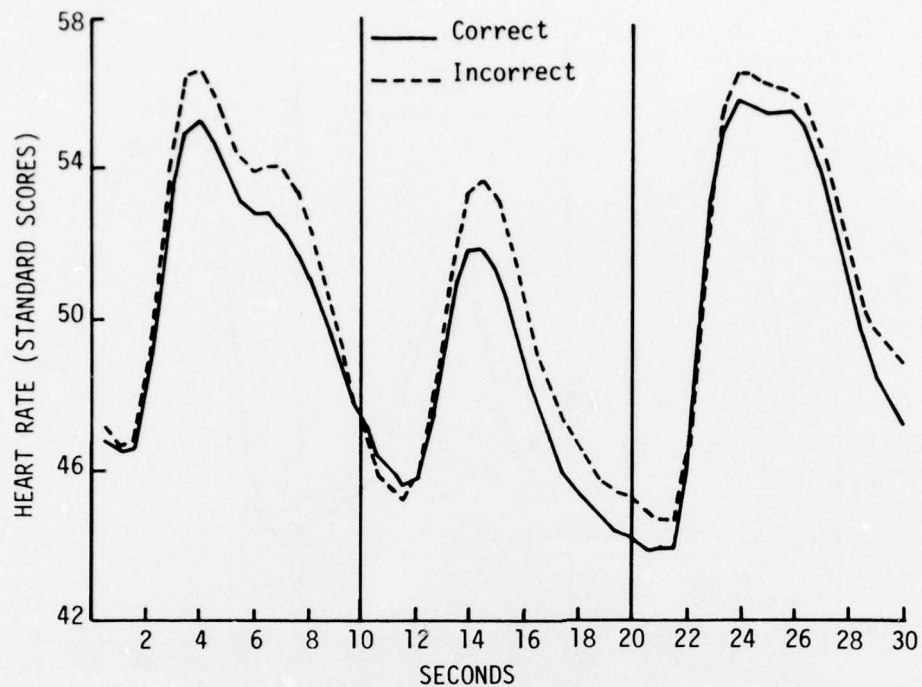


Figure 3. Averaged response curves for HR STIM(ALL)-SS.

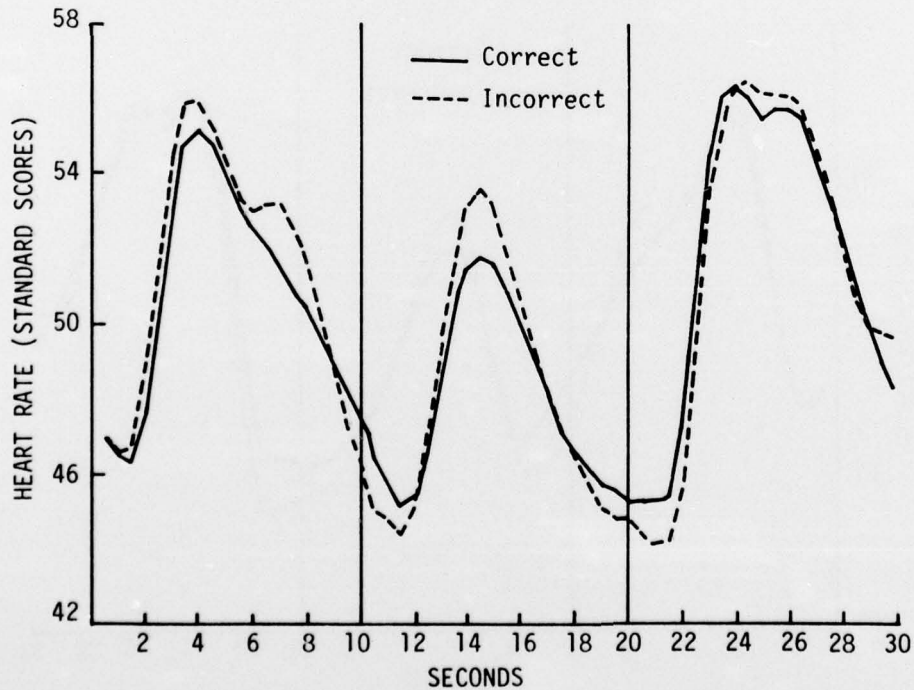


Figure 4. Averaged response curves for HR STIM(3,4)-SS.

one other occurrence, of substantial duration, where significant differences are found. In the confidence phase of the STIM(ALL)-SS analyses, the accelerative peak for the incorrect curve is significantly greater than that for the correct curve in terms of the critical-difference analysis even though the AxT interaction was non-significant.

These last two results--greater accelerative activity during incorrect performance in the confidence phase--may reflect some form of internal activity such as daydreaming that would contribute to the subsequent poorer performance during the stimulus phase.

Three significant AxT interactions were found in the RESP analyses. These all occurred during the response phase and are of uniformly brief durations as indicated by the critical-difference analysis. In the one response phase where this interaction is nonsignificant (STIM[ALL]-SS), it is just barely so, $p < .08$. The critical-difference analysis consistently places the location of this interaction in the vicinity of the 21.5th second--the point of maximum inspiration. Since the correct curve exceeds the incorrect curve at this point, the results suggest that deeper concentration during correct performance in the stimulus period may have reduced the S's oxygen intake and led to this compensatory action. (See Figures 5 through 8.)

In the stimulus phase, the analysis of EMG activity produced three significant AxT interactions with the fourth just missing significance, $p < .08$. The critical-difference analysis of STIM(ALL)-RS, STIM(ALL)-SS, and STIM(3,4)-SS data consistently places the location of this interaction at the last few seconds of this phase. Contrary to expectations from activation theory, greater muscle tension is found during incorrect performance. It is possible that what is being exhibited by this difference is greater restlessness or that a level of activation has been reached that is incompatible with optimal performance on this task. The anomalous result obtained during the STIM(3,4)-RS stimulus phase, where significantly greater EMG activity was found for correct performance during the early seconds of this period, is difficult to interpret and may simply reflect a sampling oddity. (See Figures 9 through 12.)

A significant interaction was also found in the confidence phase of STIM(ALL)-RS analysis and, together with the other significant differences exhibited by the critical difference analysis, displays the same relationship found in the stimulus phase--greater EMG activity during incorrect performance--and may be accounted for in the same terms as above.

The differences at all points in time between the correct and incorrect SCL curves, with the exception of two (the 9th and 9.5th second) in the confidence phase of STIM(3,4)-SS data, exceed the

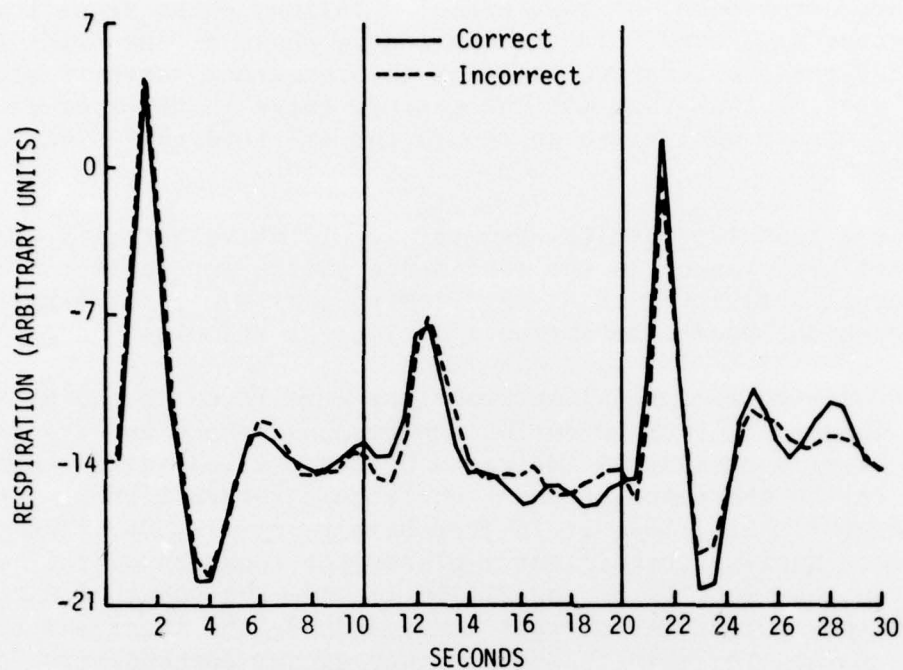


Figure 5. Averaged response curves for RESP STIM(ALL)-RS.

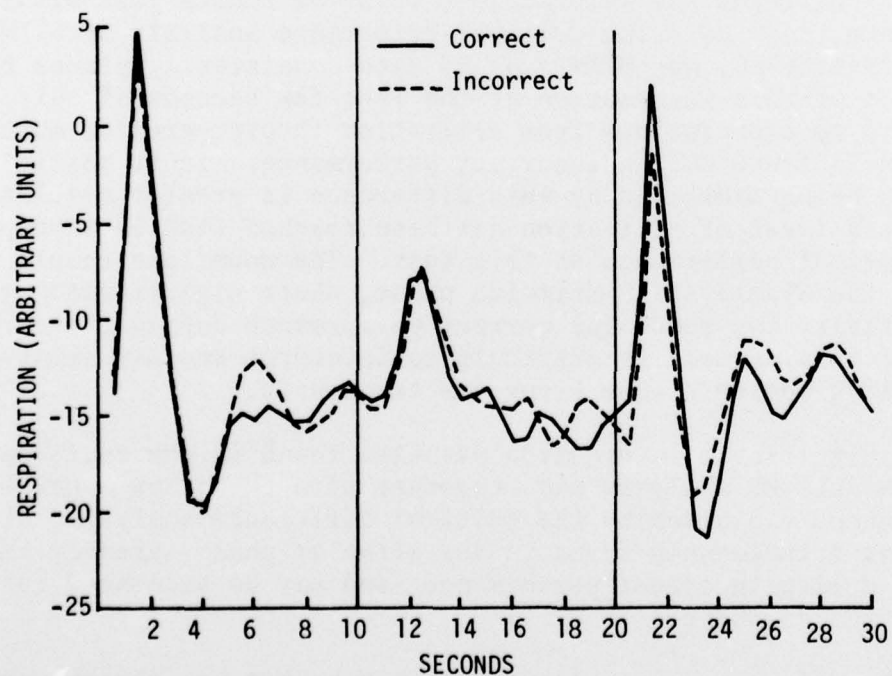


Figure 6. Averaged response curves for RESP STIM(3,4)-RS.

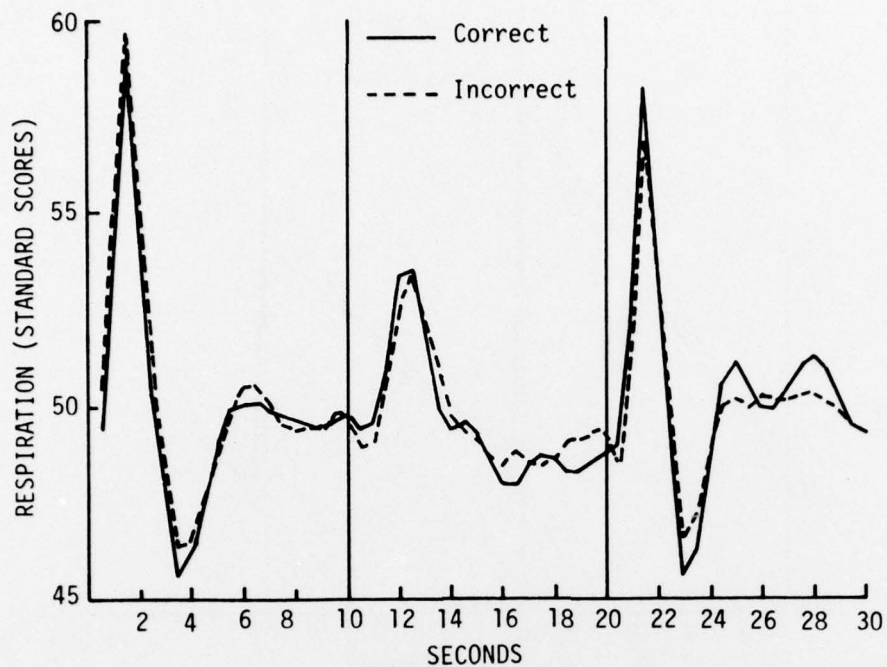


Figure 7. Averaged response curves for RESP STIM(ALL)-SS.

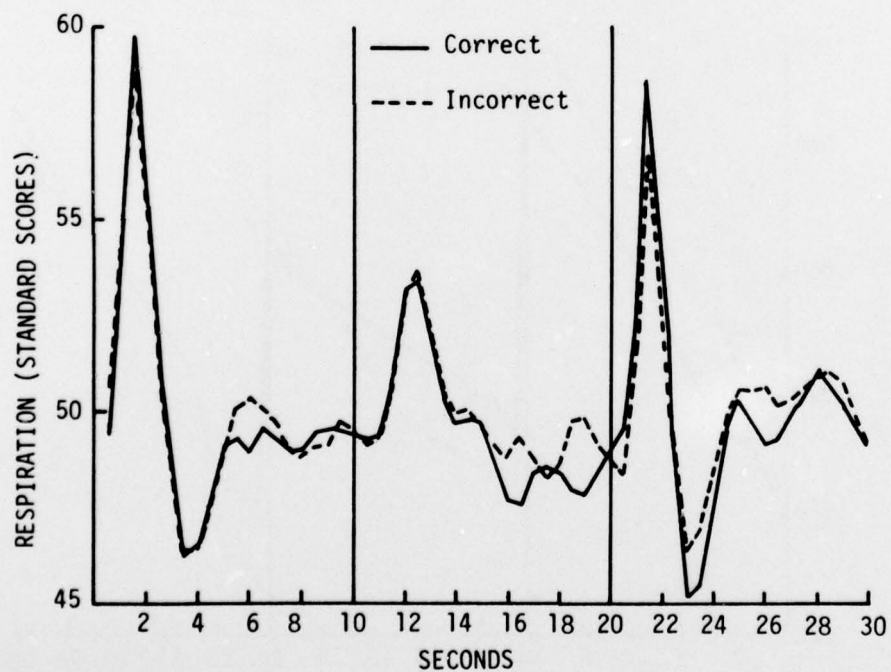


Figure 8. Averaged response curves for RESP STIM(3,4)-SS.

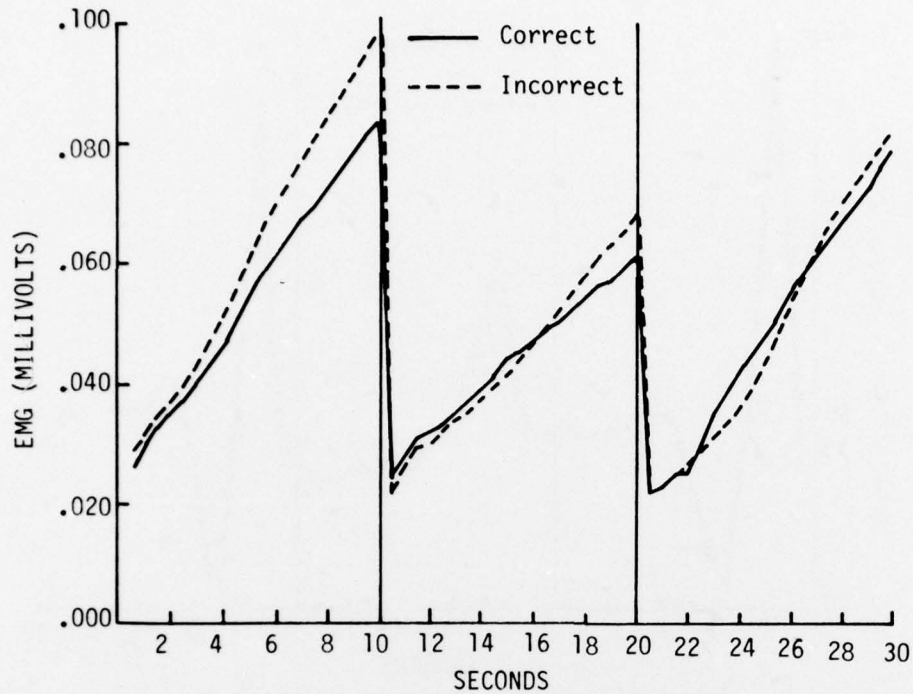


Figure 9. Averaged response curves for EMG STIM(ALL)-RS.

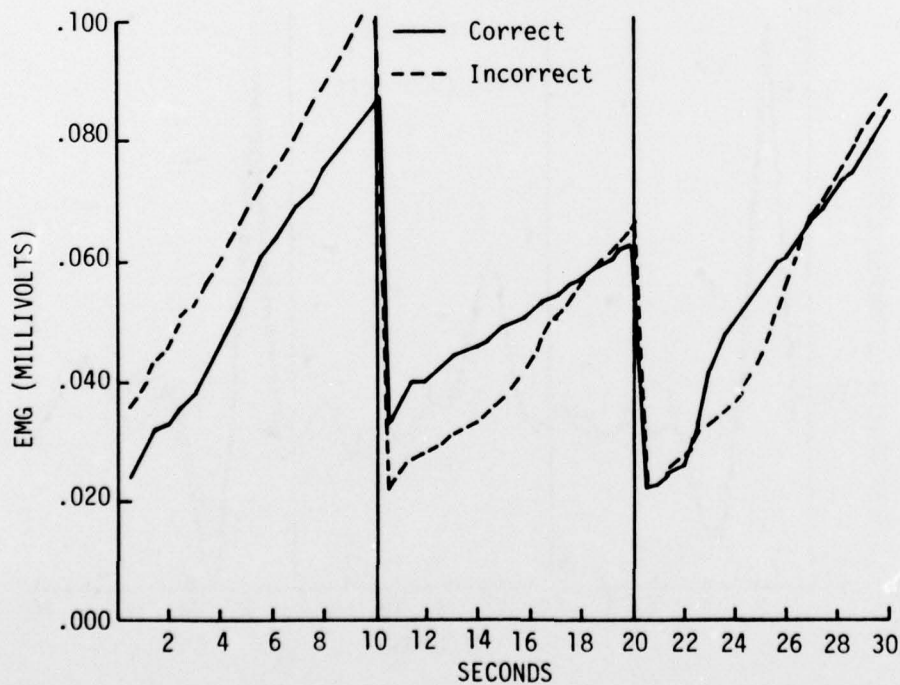


Figure 10. Averaged response curves for EMG STIM(3,4)-RS.

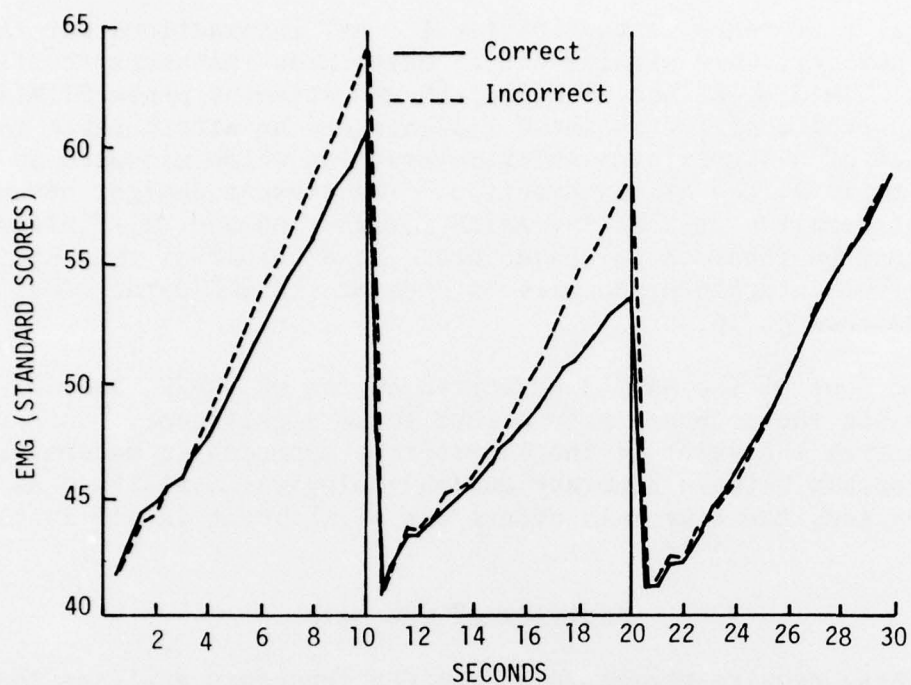


Figure 11. Averaged response curves for EMG STIM(ALL)-SS.

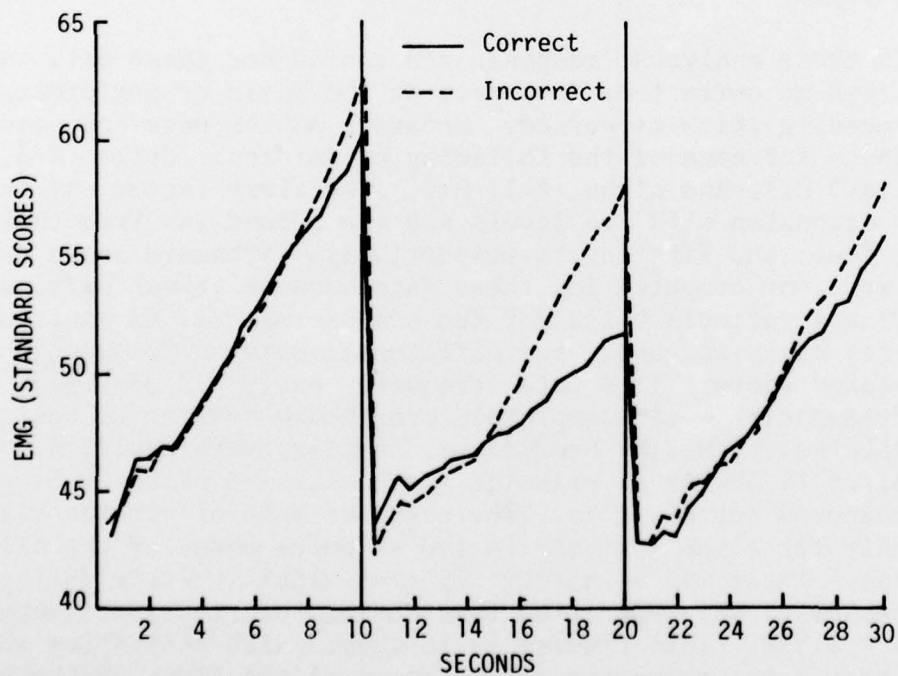


Figure 12. Averaged response curves for EMG STIM(3,4)-SS.

critical difference at the .05 level. AxT interactions for these data, however, were significant in only three instances--confidence phases STIM(3,4)-RS and STIM(3,4)-SS and stimulus phase STIM(ALL)-SS. The apparent conflict in these findings may be attributable to the presence of a significant AxTxS interaction which was used as the error term for the AxT interaction. The present design, however, did not permit a test of the AxTxS interaction and this interpretation must be regarded as conjecture. As a result of this ambiguity, no further interpretation will be made of the SCL data. (See Figures 13 through 16.)

In none of the ANOVAs performed on the HR, RESP, and EMG variables was the accuracy main effect found significant. This result underscores the value of the interactive approach in determining the relationship between accuracy and physiological activity. As would be expected, the time main effect was significant in all instances.

Analyses of EEG Data

These results are based on the EEG frequency analyses that were performed in real time and outputted to magnetic tape. The amount of data to be stored from the other parameters precluded the storage of the raw EEG signal which was digitized at approximately 100/second and resulted in the accumulation of 1024 data points in a 10-second period. As a consequence, averaged EEG response curves could not be developed.

In these analyses, response and confidence phase data were classified as correct or incorrect on the basis of performance in the preceding stimulus period. Separate ANOVAs were computed for each phase for each of the following bandwidths: delta (1-3 Hz), theta (4-7 Hz), and alpha (8-12 Hz). The first factor was the accuracy dimension with two levels and the second was frequencies with three, four, and five levels respectively. Standard score transforms were not computed for these data because it was difficult to determine a suitable basis for the transforms. If an individual frequency basis was used, the differences between frequencies would be obscured whereas if a total frequency basis was employed, dominant frequencies would completely overshadow smaller values. The STIM(ALL) vs. STIM(3,4) breakdowns, however, were analyzed. Thus, a total of 18 ANOVAs (2 stimulus treatments x 3 phases x 3 bandwidths) were computed for EEG data. The accuracy main effect was significant only for alpha activity in the stimulus phase of the STIM(ALL) analysis. There was significantly less alpha activity during correct performance in this condition than during incorrect performance, $F(1/9) = 5.500$. This finding is in accord with activation theory which associates reduction or abolition of the alpha rhythm with increased activation.

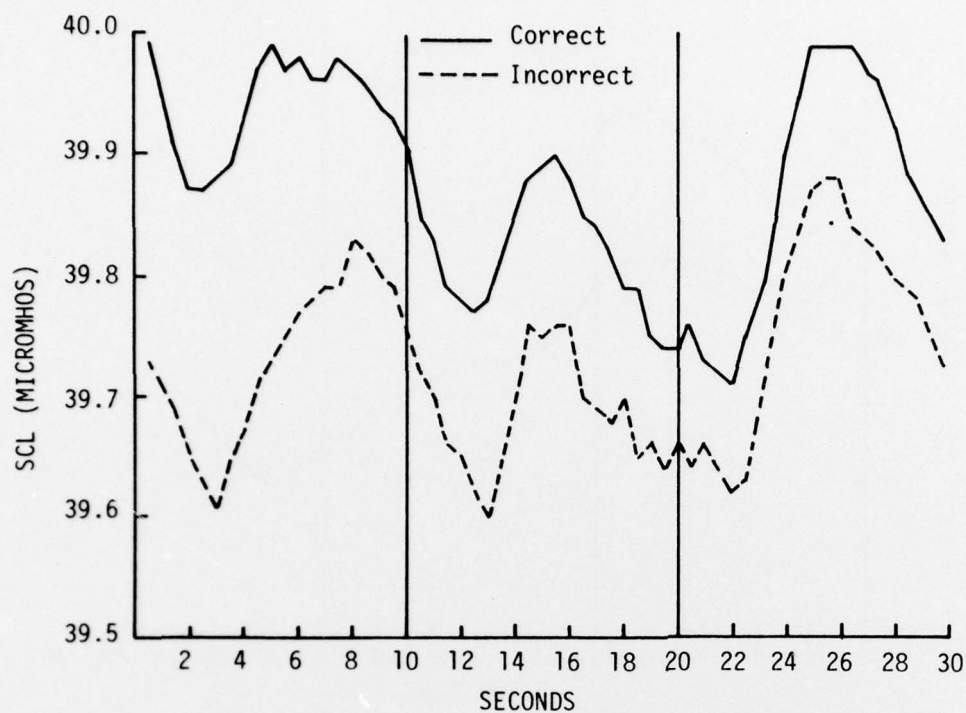


Figure 13. Averaged response curves for SCL STIM(ALL)-RS.

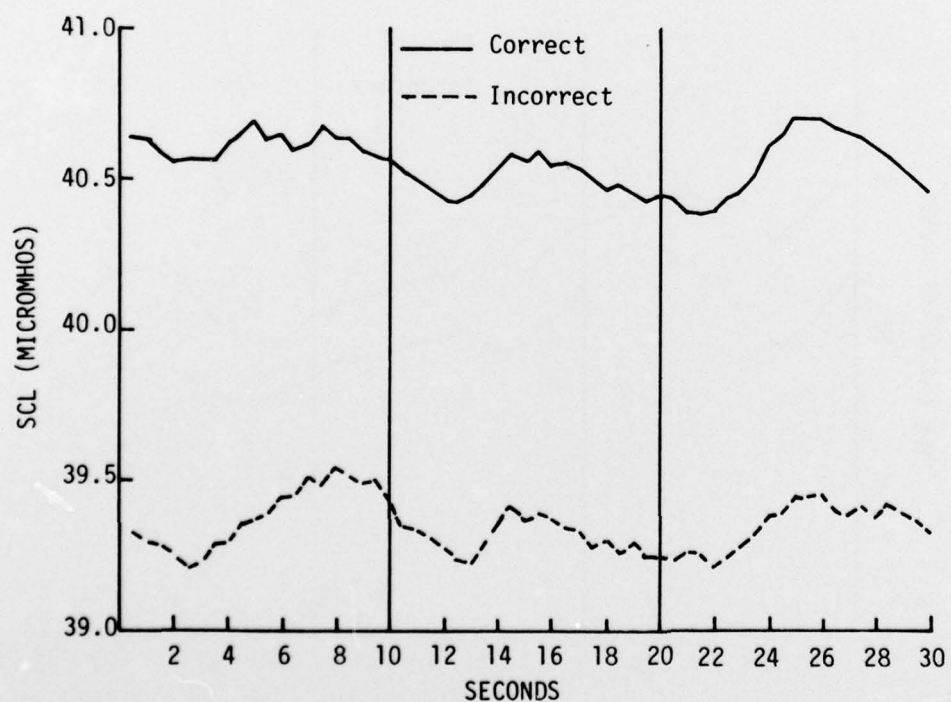


Figure 14. Averaged response curves for SCL STIM(3,4)-RS.

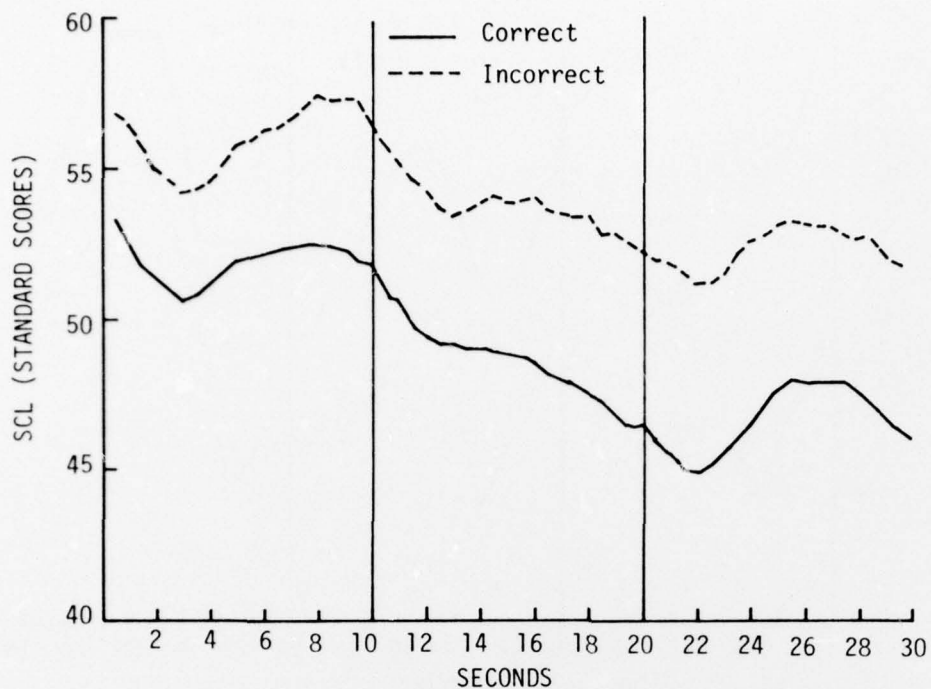


Figure 15. Averaged response curves for SCL STIM(ALL)-SS.

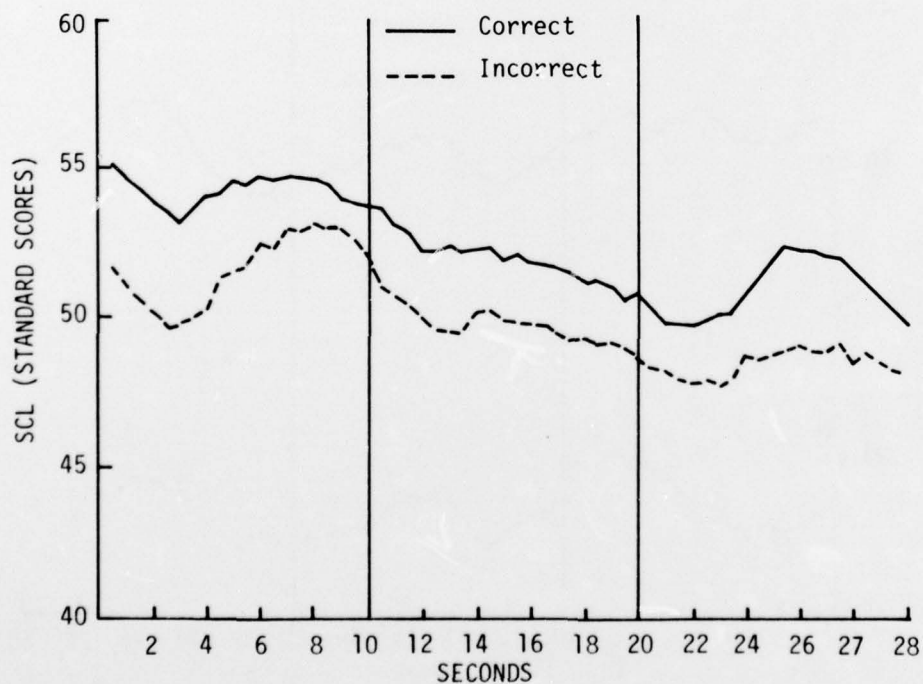


Figure 16. Averaged response curves for SCL STIM(3,4)-SS.

All of the frequency main effects proved significant with the exception of alpha activity in the stimulus phase of both the STIM-(ALL) and STIM(3,4) data treatments. None of the accuracy x frequency interactions were statistically significant. The lack of interaction between the frequency components of a bandwidth and accuracy indicates that it would not be profitable to examine individual frequencies as discriminators of correct from incorrect performance.

DISCUSSION

The evidence from this experiment illustrates the fruitfulness of the interactive approach in assessing the relationship between performance and physiological variables. Not one significant F-value was obtained for the accuracy main effect in the ANOVAs performed on the HR, RESP, and EMG data, a result that would have led to rejection of these variables as effective discriminators. The interaction of accuracy with time of measurement, on the other hand, proved significant in many instances and, more important, the pattern of the significant interactions was meaningful in terms of both known and hypothesized relationships between physiological activity and performance.

In addition, the results indicate that when the operation of task difficulty is eliminated or diminished (the STIM[3,4] treatment) the relationship between accuracy and performance does not appear (at least not at a statistically significant level) when analyses are based on raw score data. The desired statistical sensitivity reappears, however, when analyses are based on standard score transforms of the data. These findings are particularly important since in the former instance (difficulty operative) it is likely that the differential difficulty is producing the physiological changes, whereas in the latter case (difficulty eliminated) it is more likely that the physiological changes are affecting adequacy of performance. Since in monitoring and vigilance tasks we are concerned with predicting when an operator is likely to start missing signals, it is more essential that we be able to use the physiological changes to predict performance changes rather than be able to use task demands to predict physiological changes. The standard score transform appears to be a suitable method of achieving this goal.

The results taken as a whole, coupled with recent improvements in computer technology and physiological monitoring equipment, suggest that it should be possible to construct a practical alertness indicator. It would be feasible, for example, to determine optimal physiological response curves for an operator under controlled conditions, store this profile on a mass storage device, and then test on-going physiological activity during actual working conditions against this profile for critical deviations from acceptable levels.

NOTE

The author gratefully acknowledges the assistance of John Fite, Jr., in the data collection phase of this experiment and the heroic efforts of Joan Craig and Michael Cameron in battling a balky computer to produce the extensive analyses that appear in this report.